Weekly Report

Period: 2016/7/18-2016/7/24
Reporter: Li Zongzhuang

Knowledge Graph Visualization for Understanding Ideas **(from IJCDSE 2013 ,Author** Allen Zhu)

This paper discusses the potential of knowledge graphs.

The author provides an algorithm to visualize knowledge graphs. And they have shown the explortion of entire graphs through Desic framework which is made by Javascript and HTML5 canvas.

This paper uses Self-Organizing Maps.

Algorithm 3: Desic algorithm for graph layout

input: Graph G
output: layout of G on the sphere

$r :=$ diameter(G)
$iterations := [1000/r]$
$\lambda := 10000$

while $r > 0$:
    generate a random vector $\vec{v}$
    find node n with $min(\|\vec{v} - \vec{n}\| - B_n)$
    for all nodes n_i with $dist(n, n_i) \geq r$:
        $\overrightarrow{n_i} = normalized[\overrightarrow{n_i} + h(dist)\alpha(\vec{v} - \overrightarrow{n_i})]$

    if($iterations = 0$)
        decrement r
        iterations $= [1000/r]$
    else decrement iterations

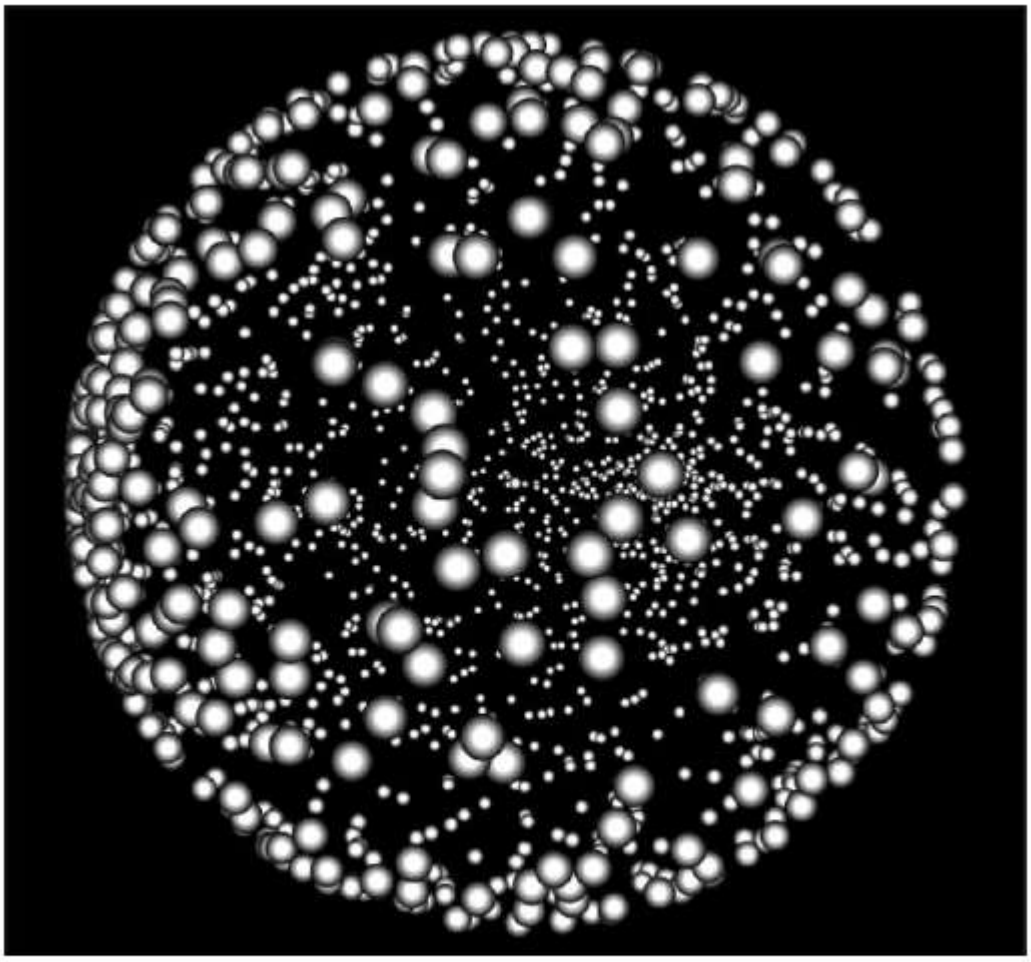
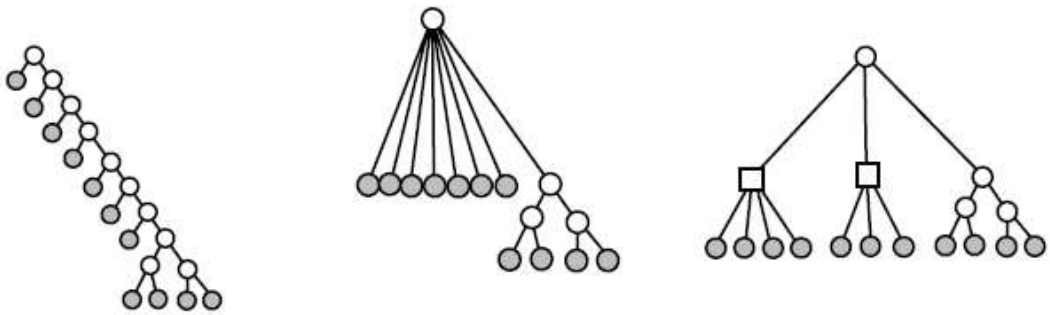


Figure 3. Desic visualization of the full 1718-node *Connections* graph, rendered in-browser (edges not shown)

ASK-GraphView : A Large Scale Graph Visualization System **(from TVCG 2006, Author** James Abello and so on)

ASK-GraphView, a node-link-based graph visualization system that allows clustering and interactive navigation of large graphs, ranging in size up to 16 million edges.

By lowering the interactivity requirements it can scale to substantially bigger graphs. The user is allowed to navigate this hierarchy in a top down manner by interactively expanding individual clusters. ASK-GraphView also provides facilities for filtering and coloring, annotation and cluster labeling.



(a) (b) (c)

The changing of data structure.

附录：

Visual Exploration and Analysis of Knowledge Graph

Li Zongzhuang

Abstract: The Knowledge Graph is a knowledge base used by Google to enhance its search engine's search results with semantic-search information gathered from a wide variety of sources. The Knowledge Graph uses the power of semantics, and wants to improve search precision and effectiveness by building the semantic web. Visualization is the study of (interactive) visual representations of abstract data to reinforce human cognition. The visualization of Knowledge Graph can Effectively improve the efficiency of the user to complete the search target and precision. Data integration and correlation analysis reasoning is one of the best visual analysis applications. At this time, There are many applications have been exploited based on this concept.

1. Introduction

Knowledge graphs on the Web are a backbone of many information systems that require access to structured knowledge. The idea of feeding intelligent systems and agents with general, formalized knowledge of the world dates back to classic Artificial Intelligence research in the 1980s. Then, with the advent of Linked Open Data sources like DBpedia, and by Google's announcement of the Google Knowledge Graph in 2012, representations of general world knowledge as graphs draw a lot of attention again.

In 2014, Google announced a new initiative,called the Knowledge Vault, which derives much of its data from the Knowledge Graph and the sources thereof, as well as harvesting its own data, ranking its reliability and compiling all results into a database of over 1.6 billion facts collected by machine learning algorithms.

In a paper , author shows what is a knowledge graph:

1. mainly describes real world entities and their interrelations, organized in a graph.
2. defines possible classes and relations of entities in a schema.
3. allows for potentially interrelating arbitrary entities with each other.
4. covers various topical domains.

Vision is the most important channels to the information of the outside world. Visualization is the data technology of interactive visual expression. On the century of big data, The ability of processing data is far behind the ability to get the data. The amount of data contained in Knowledge Graph is huge, so the visualization can be an important means of Knowledge Graph data processing. It can help us find the phenomena and laws faster and achieve the goal. However, the research about the visualization of Knowledge Graph is relatively shallow.

2. Knowledge Graph

2.1 Knowledge Graphs

There are many ways to build knowledge graphs. They can be curated like *Cyc*, edited by the crowd like *Freebase* and *Wikidata*, They can also be extracted from large-scale, semi-structured web knowledge bases such as Wikipedia, *DBpedia* and *YAGO*. Furthermore, information extraction methods for unstructured or semi-structured information are proposed, which lead to knowledge graphs like *NELL*, *PROSPERA*, or *KnowledgeVault*.

Freebase, a public, editable knowledge graph with schema templates for most kinds of possible entities. The last version of Freebase contains roughly 50 million entities and 3 billion facts. Freebase's schema comprises roughly 27,000 entity types and 38,000 relation types.[1] It have been shutdown because of company aquired by Google.

DBpedia is a community effort to extract structured information from Wikipedia and to make this information available on the Web. Keys are mapped to properties in that ontology. Based on those mappings, a knowledge graph can be extracted. It contains 6.2 million entities and 187 million statements about those entities. [2]The ontology comprises 735 classes and 2,800 relations.[]

After the shutdown of Freebase, the data contained in Freebase is subsequently moved to Wikidata.[]In Wikidata, for each axiom, it's provenance metadata can be included.[] Wikidata contains roughly 19 million instances and 100 million statements. [3]Its schema defines 23,000 types and 1,600 relations.

Google's Knowledge Graph was introduced to the public in 2012, and it was the tern knowledge graph being coined. Google's Knowledge Graph display was added to Google's search engine in 2012. Once a user search one thing, it provides structured and detailed information about the topic in addition to a list of links to other sites. According to Google, the information in the Knowledge Graph is derived from many sources, including the CIA World Factbook, Wikidata, and Wikipedia.It contains 18 billion statements about 570 million entities, with a schema of 1,500 entity types and 35,000 relation types.[]

Never-Ending Language Learning is an implementation of the Read the semi-structured Web data approach. [] As opposed to DBpedia, all facts recorded by NELL can be tracked according to its provenance and a degree of confidence.[] Nell2RDF platform can transform the data generated by NELL into state of the art Linked Data, following best practices.[] NELL has been learning to read the web 24 hours/day since January 2010, and so far has acquired a knowledge base with over 80 million confidence weighted beliefs (e.g., servedWith(tea, biscuits)).[]

Name	Instances	Facts	Types	Relations
DBpedia (English)	4,806,150	176,043,129	735	2,813
YAGO	4,595,906	25,946,870	488,469	77
Freebase	49,947,845	3,041,722,635	26,507	37,781
Wikidata	15,602,060	65,993,797	23,157	1,673
NELL	2,006,896	432,845	285	425
OpenCyc	118,499	2,413,894	45,153	18,526
Google's Knowledge Graph	570,000,000	18,000,000,000	1,500	35,000
Google's Knowledge Vault	45,000,000	271,000,000	1,100	4,469
Yahoo! Knowledge Graph	3,443,743	1,391,054,990	250	800

An overview about these knowledge graphs.

[1] http://www.freebase.com.
[2] http://wiki.dbpedia.org/services-resources/datasets/dataset-2015-10/dataset-2015-10-statistics
[3] https://tools.wmflabs.org/wikidata-todo/stats.php

2.2 Semantic Web

The core concept of Knowledge is the introduction of the semantic, which means that let the computers know the semantic judgments.

3. Visualization

3.1 Graph data visualization

Graph data is an important component in data. The visualizations about graph data are often presented by node-link graph.

3.2 High-dimensional data visualization

There are many data sets have more than one dimension. So many visualization tools have been invented to present high-dimensional data. The results got by Knowledge Graph often have many properties. That means we can get some revelation.

3.3 Another types visualization

Because of the difference of goals, there can be many visualization schemes. Maybe we can learn more from them.

4. Application softwares

Based on the theory of human-computer interaction, there are a lot of software is put forward based on the semantic web. This kind of software focus on data integration and correlation analysis. Data integration made in background automatically, and data correlation analysis mainly rely on people's reasoning ability as well as front end some interactions. That’s the best application, which give full play to the people the calculation of analytical reasoning skills and computer specialty.

There are many applications in this area, such palantir, IBM i2, Tableau and so on.